

Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

COMMENT

Tel : 81-45-503-9222
Fax : 81-45-503-9216

Email : genome-res@sc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N.,
.M., Kono,H., Okazaki,Y., Muramatsu,M., and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
waqi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Wataniki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A., and
Hayashizaki,Y.

RIKEN integrated sequence analysis (RISA) system -384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)

Kono,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
,Y. and Hayashizaki,Y.

Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Please visit our web site (<http://genome-gsc.riken.go.jp>) for
further details.

e mouse tissues.

Location/Qualifiers

l..622

/organism="Mus musculus"

/db_xref="taxon:10090"

/clone="G431005.l12"

/clone.lib="RIKEN full-length enriched, bone marrow stroma
cell CRL-2028 SR-4987 CDNA"
tissue_type="bone marrow"
cell_line="stroma cell"
/cell_line="CRL-2028 SR-4987"

201 a 80 c 117 g 224 t

ch 2.6%; Score 41.8; DB 9; Length 622;
1 Similarity 63.4%; Pred. No. 26;
64 Conservative 0; Mismatches 37; Indels 0; Gaps 0;

catttcacatbaaacctaagtctttccctcataatgcatttcgtgaactctcatagtg 1569
||||| |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
ATTTTACTTGTAATAAGCGCTGTATTAAAAAACCACATGGTTTTATTATATTTTG 568

cattccaaggagatatggttgtgaagaacaacaaaaaaa 1610
||||| |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
AAATTTAAAATAAATAAGCTTTAATATAAAAAAAAAAAAA 609

CNSOOLIV 1064 bp DNA linear GSS 03-JUN-1999
Drosophila melanogaster genome survey sequence T7 end of BAC #
BACH04C22 of RPCR-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
AL061101
GI:4939564
GSS.

fruit fly.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1064)
Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE [E-mail : seqreg@genoscope.cns.fr
Web : www.genoscope.cnc.fr]
determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).

Seq primer: -40RP from Gibco
POLYA-No.

FEATURES	Location/Qualifiers
source	1. .356

```

/db_xref="taxon:10090"
/clone="IMAG508"
/clone_lib="Soares mouse 3NbMS"
/sex="male"
/tissue_type="Spleen"
/dev_stage="4 weeks"
lab_host="DH10B"

```

BASE COUNT 143 a 43 c 64 g 106 t

	RESULT	14
	CNS015QP/c	
	LOCUS	
	DEFINITION	CNS015QP 739 bp DNA linear GSS 26-JUL-1999
	ACCESSION	Drosophila melanogaster genome survey sequence T7 end of BAC
	VERSION	BACN1J12 of DrosBAC library from Drosophila melanogaster (fruit
	KEYWORDS	fly), genomic survey sequence.
	SOURCE	AL105787
	ORGANISM	AL105787.1 GI:5618327
		GSS.
		fruit fly.
		Drosophila melanogaster
		Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
		Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
		Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
	REFERENCE	1 (bases 1 to 739)
	AUTHORS	Genoscope.
	TITLE	Direct Submission
	JOURNAL	Submitted (23-JUN-1999) Genoscope - Centre National de Sequencage ;
		BP 191 91006 EVR cedex - FRANCE (E-mail : seqrefgenoscope.cns.fr
		- Web : www.genoscope.cns.fr)
	COMMENT	Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk/ . This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billard at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC

project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelOBAC11.

FEATURES	SOURCE	Location/Qualifiers
	1..739	
	/organism="Drosophila	melanogaster"
	/plasmid="pbeloBAC11"	
	/db_xref="taxon:7227"	
	/clone_11b="DrosBAC"	
	/clone="BACN14J12"	
	/note="end : T7"	
BASE COUNT	130 a 153 c 117 g	229 t . 110 others
ORIGIN		

Query Match	2.6%	Score 41.6;	DB 12;	length 739;
Best Local Similarity	37.0%;	Pred. No. 29;		
Matches 80; Conservative	37;	Mismatches 99;	Indels 0;	Gaps 0

[illegible]

LOCUS	BH034649	806 bp	DNA	linear	GSS 17-JUL-2001			
DEFINITION	RPCI-24-294M13.TU RPCI-24 Mus musculus genomic clone RPCI-24-294M13							
ACCESSION	, DNA sequence.							
VERSION	BH034649							
KEYWORDS	BH034649.1	GI:14807705						
SOURCE	GSS.							
ORGANISM	house mouse.							
	Mus musculus							
REFERENCE	Euartrypa, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;							
	Mammalia; Eutheria; Rodentia; Scurionathii; Muridae; Murinae; Mus.							
AUTHORS	1 (bases 1 to 806)							
	Zhao,S., Niemman,W., Malek,J., Shatsman,S., Akintat,B., Levins,M.,							
TITLE	Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,							
	Russell,D., de Jong,P. and Fraser,C.M.							
JOURNAL	Mouse BAC End Sequences from Library RPCI-24							
	unpublished (1999)							
COMMENT	Other_GSSs: RPCI-24-294M13.TU							

/strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCI-24-294M13"
 /clone_id="RPCI-24"
 /sex="Male"
 /cell_type="Spleen/Brain"
 /note="Vector: pTARBAC1; Site 1: BamHI; Site 2: BamHI;
 RPCI-24 Mouse BAC library produced by Pieter de Jong. The
 library was cloned in the pTARBAC1 cloning vector at the
 BamHI sites using MboI partially digested male C57BL/6J
 DNA."

BASE COUNT 222 a 148 c 164 g 272 t
 ORIGIN

Query Match 2.6%; Score 41.6; DB 12; Length 806;
 Best Local Similarity 53.0%; Pred. No. 29;
 Matches 89; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

```

QY 857 ttcataatgattccctgaacaagcattgtagtgcagttatcatccaatttcgct 916
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 605 TTTTAAATTTAAACATAGACACACTGTGATAGCTGTTTAAATGCGATTGAACCT 546

QY 917 ctgaactagcaagaacatattaattacacgaatacttggagttccagatgtccccc 976
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 545 CTGAATATATTATTTAAATTAATAAATAAGGCCCACTTCTGATACCTATTATTACT 486

QY 977 gtccattggaacccgagatacaaatltacaatagttacttgaata 1024
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 485 CTTAAACTGCATAACACTACACAGTATCAGAGTGATGATCTGTATA 438
  
```

Search completed: June 12, 2002, 07:56:52
 Job time: 24667 sec

Description

1	1602.2	99.5	1630	3	AB015878	AB015878	Caribbea
2	277	17.2	1602	3	AB036714	AB036714	Carlydrea
3	48.6	3.0	7218	6	166494	166494	Sequence, 14
4	46.2	2.9	1938	6	AB045319	AB045319	Chiropsal
5	46.2	2.9	56153	6	AX251552	AX251552	Sequence
6	43.8	2.7	15548	6	AX347057	AX347057	Sequence
7	43.2	2.7	9733	6	AX345584	AX345584	Sequence
8	42.8	2.7	1141	6	AX083744	AX083744	Sequence
9	42.8	2.7	1240	3	AF104650	AF104650	Sequence
10	42.8	2.7	7001	6	AX347354	AX347354	Lastiologis
11	42.8	2.7	7002	6	AX349075	AX349075	Sequence
12	42.8	2.7	7301	6	AX345613	AX345613	Sequence
13	42.8	2.7	174019	9	AP001538	AP001538	Sequence
14	42.8	2.7	340000	9	AP001674	AP001674	Homo sapi
15	42.4	2.6	195721	2	AC027059	AC027059	Homo sapi
16	42.2	2.6	96975	2	AC004975	AC004975	Homo sapi
17	42.2	2.6	349980	6	AX344564	AX344564	Sequence
18	41.8	2.6	61143	2	AC099802	AC099802	Homo sapi
19	41.8	2.6	200317	9	HS69813	HS69813	Human DNA
20	41.6	2.6	254534	2	AC097326	AC097326	Pan trogl
21	41.6	2.6	140934	2	AC009949	AC009949	Homo sapi
22	41.6	2.6	169728	2	AC026183	AC026183	Homo sapi
23	41.4	2.6	349980	6	AX344565	AX344565	Sequence
24	41.2	2.6	38342	6	AX251504	AX251504	Sequence
25	41.2	2.6	38342	6	AX344503	AX344503	Sequence
26	41.2	2.6	127197	6	AX196294	AX196294	Sequence
27	40.8	2.5	87487	10	MMU427334	MMU427334	Mus muscu
28	40.8	2.5	5216	6	AX281348	AX281348	Mus muscu
29	40.6	2.5	5216	6	AX345249	AX345249	Sequence
30	40.6	2.5	7113	6	AX345707	AX345707	Sequence
31	40.6	2.5	183272	2	AC094428	AC094428	Rattus no
32	40.6	2.5	205472	2	AC005506	AC005506	Plasmodiu
33	40.6	2.5	256172	2	AC005139	AC005139	Plasmodiu
34	40.6	2.5	310779	2	AC005140	AC005140	Plasmodiu
35	40.4	2.5	7037	6	AX346653	AX346653	Sequence
36	40.4	2.5	7195	6	AX277866	AX277866	Sequence
37	40.4	2.5	7195	6	AX323551	AX323551	Sequence
38	40.4	2.5	158660	2	AC068425	AC068425	Mus muscu
39	40.4	2.5	188826	9	AL359380	AL359380	Human DNA
40	40.2	2.5	3678	10	BC011531	BC011531	Mus muscu
41	40.2	2.5	303750	1	AP003133	AP003133	Staphyloc
42	40.2	2.5	346900	1	AP003362	AP003362	Staphyloc
43	40	2.5	47703	3	U23511	U23511	Caenorhabd
44	40	2.5	106706	6	AC008465	AC008465	Homo sapi
45	40	2.5	152771	9	AP004241	AP004241	Homo sapi

ALIGNMENTS

Accession	Length	Type	Version
AB015878	1630 bp	mRNA	linear
Caribbea rastonii crt-1	mRNA for toxin-1, complete cds.	INV	06-SEP-2000

AB015878.1 GI:9988401

Carybdea rastonii cDNA to mRNA

Eukaryota; Metazoa; Cnidaria; Cubozoa; Cubomedusae; Carybdeidae;

1 (sites)

Nagai, H., Takuwa, K., Nakao, M., Ito, E., Miyake, M., Noda, M. and Nakajima, T.

Novel proteinaceous toxins from the box jellyfish (Sea wasp
carybdea rastonii)

Biochem. Biophys. Res. Commun. 275 (2), 582-588 (2000)

20422301
2 (bases 1 to 1630)
Nucleotide sequence

Submitted: 20 May 1999

Submitted (29-JUN-1998) HIROSHI Nagai, Suntory Institute for Bioorganic Research; Wakayamadai 1-1-1, Shlmamoto, Osaka 618-8503

Japan (E-mail: Hiroshi_Nagai@suntoory.co.jp, Tel: 81-75-962-3743,
Fax: 81-75-962-2115)
Location/Qualifiers

FEATURES

source

1..1630
/organism="Carybdea rastoni"
/db_xref="taxon:78582"

gene

28..1380
/gene="crt-1"

CDS

/codon_start=1
/product="Toxin-1"
/protein_id="BAB12728.1"
/db_xref="gi:9988402"
/translation="MILKHPMLFVLAITSKHKGRSDVNSLTKVETALKEASGN
EALFALGKGEIOFKPDVKGATKILGSGASDGRKTIISGCDIYAGIA
TTTGGPVKMGIGAVASVYSLISFTGSAKNSAAVYIDRLSKRDEAIDRHAGAK
RDEASSAFIDVKKQSNLTDSDLSITANVPVKFSEFISQLESRIISGAAATSLSD
AKRAVDILLYCOLVYMRRETLVDLAIIYRRGNMEHVAENARVKELEAATLDF
LHKLIPEOALIGAVYHPIASSETSKAILNTYKFEVPPVPRIGNRKFTNSWNTY
STCSAYMGNTWFRGCSNVNRPNTFVSKMSGDFVEMNSDRKLYITKHDOGMGCTL
DEPDGDQHMRIPILRHGYMYSKRWRNWEYMESSASGIIIRSWENNPFGQHWSTI
"

BASE COUNT 517 a 314 c 354 g 445 t
ORIGIN

Query Match 99.5%; Score 1602.2; DB 3; Length 1630;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1604; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Oy 1 gcaacagcagctcgtggaagagagaccatgatctgaacatcttcctgctcttatt 60
Db 1 GCACAAGCAGCTTGTGTAAGAGACACCATGATTTCTAACAATCTTCTGGCTTTATT 60
Oy 61 gtccctgcaattactctcgcacaacacatggcaaacgctcgtgltcaatcttactt 120
Db 61 GTCCCTGCAATTAATCTCTGCAAAACATGCAAAACGCTGATGTCATTTCTTACTTACT 120
Oy 121 aaggtgaagacgtccttaaaagaagcttcgtgtaagaagagctgctcttgaagctt 180
Db 121 AAGGTGAAGACGTCTTAAAGAGAGCTTCTGTGAGCAAGAGGCTCTTGTAGGCTTTTA 180
Oy 181 gaaggtctaaagaagagatccagacaaaacagaccgagcttggacaagaacccaaaatc 240
Db 181 GAAGGCTTAAAGAGAGAGATCCAGACAAAACAGACGAGTTGACAAAGCAACAAAATTC 240
Oy 241 ctgtgacatcgtcgagatcagcttaagaataaattctggagatggcaacaaatcatt 300
Db 241 CTGTGACATCTGTCGATCAGCTTAGAATAATTAAATTTGAGATGCAACAAAATCATTT 300
Oy 301 tctggtgctgctgacattgtgtcagaagaattgcaacaacttttggagccctgtcgagatg 360
Db 301 TCTGTTGCTCTGCACATTTGTTGAGAGAAATTTTGGAGGCCCTGTGCGGAGTG 360
Oy 361 ggaatcgagcgttaagcttcttcttctcaatcctatcattgttcttgaagctca 420
Db 361 GGAATCGAGCGTAACTTCTTTTGTTCATATTCATATGTTTACTGGAACCTCA 420
Oy 421 gcaaaagactcagttcgtccgtatattgataagctttaaagaagaatcgagatgagcc 480
Db 421 GCAAAAGACTCAGTTGCTGCGTTATTGATAGAGCTTTAAGCAAGCATGCGATGAGGCC 480
Oy 481 atccaagaacatgacagcaggtgcacaagaagaatttgcgtgaatcatcgtcatcattcag 540
Db 481 ATCCAAGAACATGACAGCAGGTGCCAAGAGAGATTTCCTAATCATCTGCATTCATTTCAG 540
Oy 541 gtcaatgaacagcagtcacacttcaacagatagcgaactaagtaactatgagcgaatgatt 600
Db 541 GTCAATGAACAGCAGTCACACTTCAACAGATAGGACCTTAAGTATCATTTGACGGAATGTT 600
Oy 601 cctgttataaatcttgaatttatctgcagcagttgagagcagaatttcccaagcagca 660
Db 601 CCTGTTATAAATCTTGAATTTATCGACAGCTGTGAGAGACAGAAATTTCCCAAGCGCA 660
```

```
Db 601 CCTGTTATAAATTTGAATTTATCGACAGCTGTGAGAGACAGAAATTTCCCAAGCGCA 660
Oy 661 gcaactaccagctcttagcagatgcaagaagagcgttgaactcatctcgtctattgcaa 720
Db 661 GCAACTACCAGCTTtagcagatgcaagaagagcgttgaactcatctcgtctattgcaa 720
Oy 721 ctgttagcagatgcaagaagaaccttgcgtgcagcttgcgtatcttctcaagaagaagaat 780
Db 721 CTGTtagcagatgcaagaagaaccttgcgtgcagcttgcgtatcttctcaagaagaagaat 780
Oy 781 gcagaacacgtgagcaagtgctgtgtgaaacgctaaatagggtlaaacaagaagcagctgct 840
Db 781 GCAGAACACGTGAGCAAGTGCTGTGAAACCGCTAATAGGGTAAACAAAGAGCTAGCTGCT 840
Oy 841 gatcccccaggtttcttcaaaaatgatcttccaaagaagattgataagtgagctttat 900
Db 841 GATCCCCAGGTTTCTTCAATTAATGATTCCTAACAAGCATTTGATGAGCGCAGTTTAT 900
Oy 901 catccaattctgcctcgtgaacacagcaacaatatlaactacaogaatatcttggga 960
Db 901 CATCCAATTTCTGCCTCTGGAACCTAGCAACGCAATATTAAATTACAGGAATACTTTGGA 960
Oy 961 gtccagatgttcccgctccatttggaaacccgacagatacaaatlacaatatgttactgg 1020
Db 961 GTTCCAGATGTTCCCGCTCTATTGGAACCGCAGATACAAATTTACAAATAGTTACTGG 1020
Oy 1021 aatacctacagtatatgagtgagtgagcttacaatgggaattacatgcttcaagagctgtct 1080
Db 1021 AATACCTACAGTATATGATGAGTGAGGCTTACATGGAAATTCATGTCAGAGGCTGTCT 1080
Oy 1081 aacgttcggaatcccaatcaggtgacccaagaatgtctgtgtgtgttttacaacatggag 1140
Db 1081 AACGTTCCGAATCCCAATATCAGGATATCCAAATGTCTGATGGGTTTACACCATATGAG 1140
Oy 1141 aatagcagatcgaggaagtgatatacaaccaagaatgagcaaggaatggagtgagtgact 1200
Db 1141 AATAGCAGATCGGAGGAATGATATATACCAAGCATGACCAAGATGGGATGGGATGACT 1200
Oy 1201 ttgagtagagatccaggtgagcaagaagcacaatagagtgatcttcttggagcagtggaag 1260
Db 1201 TTGAGTAGAGATCCAGGTGAGCAAGGCCATATGAGTTTCAATCTTGTGAGCATGGGAG 1260
Oy 1261 tatalgtgaagctcctaagaagtgagcccaactggttcatgtlatatgaatacagtgccagt 1320
Db 1261 TATATGTGAAGCTTAAGAGAGTGCCCAACTGTTGATGATATGATGATCAAGTGCCAGT 1320
Oy 1321 ggtcactatcgtcagctgaggaataatccaggaacctcaaggaacttggagtatacataa 1380
Db 1321 GGTCACTATTCGACGTGGGAAATTAATCCAGGACCTCAAGGACATTTGGAGTATAACATA 1380
Oy 1381 ttaagaaggaatcaacaatgctccaaaggaatagaataataagaacatcaagaatgag 1440
Db 1381 TTAAGAAGGAATCAACAATGTCCTCAAAAGCATATGAAATATGAACATCAAGATGCGAG 1440
Oy 1441 tacttaagatgcacactgtatcttcatatagagatgtgtcatbaaagtcataaacat 1500
Db 1441 TACTTAAGATGCACACTGTATTTCTACATAGAGATGTCGTATGAAGTCATTAACCAT 1500
Oy 1501 ccagcggacataattcatatlaaacataatgcttccctlaataagcatttcatatgaatc 1560
Db 1501 CCAGCGGACTAATTCAATTAATAACATTAATGTTCTTAAATGAAGCATTTTCATGAATATC 1560
Oy 1561 tctatttgacatctcaagaagatattgttgaagaagaacaaaaa 1607
Db 1561 TCTATTGTGACATTTCAAGAGGATATGTTTGAAGAAACAAATTCAAA 1607
```

RESULT 2
AB036714 LOCUS AB036714 1602 bp mRNA linear INV 06-SEP-2000
DEFINITION Carybdea alata cat-1 mRNA for toxin-A, complete cds.
ACCESSION AB036714
VERSION AB036714.1 GI:9988403

KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
sig-peptide
gene
CDS
BASE COUNT
ORIGIN
Query Match
Best Local Similarity
Matches 634; Conservative 0; Mismatches 525; Indels 12; Gaps 3;
17.2%; Score 277; DB 3; Length 1602;
54.1%; Pred. No. 1.8e-55;
0; Mismatches 525; Indels 12; Gaps 3;
1.1602
/organism="Carybdea alata"
/db_xref="taxon:112899"
73.126
/gene="cat-1"
73.1464
/gene="cat-1"
73.1464
/gene="cat-1"
/codon_start=1
/product="toxin-A"
/protein_id="BAB12727.1"
/db_xref="GI:9988404"
/translation="MSRGSLSHLVLEFVISTAFPSQARLSRYRRAADAVSTIDIGIT
GOLNDGDTKRLKALOGVOEAVKKEPATTTAKSTIVGSGLSKFSQDPEPDA
SCGLDIASVATFGPGYGAIGAVASLISLSLSESGMSKSAIKOVIDAFKRYRD
OLEEDNVGAKRTFENAVITFVNSVSKTNLEVDVADVADAFNMGLVLSRI
NRGSVSTDNNEAMRTINFILYLOLSVRETLTQVILLYKRAAGAYDELAISLST
DONKEATRETVLEHOMKYSILGSSYYRPIDHSKRAIILKLTFFPGVPPAPRTFD
GLYVROMANMRYSSICKESYAGNMHFGCDSSYHGRIKLENGYHTTILRSAMT
VTKHAGKMGWGTADDDPBGQGFITFIPLNGRYMVSTKMPDYFVYMSSAHGYTRSW
HYNPPOQWKL"

581 gtaatcgcagcgaatgctcgttttaataattagtaattttagcgacagtggaga 640
659 attcgctcagcgaatgctcgttttaataattagtaattttagcgacagtggaga 718
641 gcaagaattcccaagcgagcaactaccagcttagcgatgcaagaagagcgcttagc 700
719 gcagatcaatgctcgttttaataattagtaattttagcgacagtggaga 778
701 tcatcgtcctcattgtaactgtaactgtaactgtaactgtaactgtaactgtaact 760
779 tcatcgtcctcattgtaactgtaactgtaactgtaactgtaactgtaactgtaact 838
761 tctcctcagcgaagaatgtaactgtaactgtaactgtaactgtaactgtaactgtaact 814
839 ttttgcacagcgtcgcgctgctgctgctgctgctgctgctgctgctgctgctgct 898
815 ataggtaacacaagagctagctgctgctgctgctgctgctgctgctgctgctgctg 874
899 gtgacaaacaaagagcagcagcagcagcagcagcagcagcagcagcagcagcagc 958
875 aacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 934
959 actatctcctcgtgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1018
935 tattaattacagaagaatgtaactgtaactgtaactgtaactgtaactgtaactgtaact 991
1019 ttttgcacagcgtcgcgctgctgctgctgctgctgctgctgctgctgctgctgct 1078
992 gcaagtaacacaagagctagctgctgctgctgctgctgctgctgctgctgctgctg 1051
1079 ttttgcacagcgtcgcgctgctgctgctgctgctgctgctgctgctgctgctgct 1138
1052 tgggaatcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1111
1139 cggcgaatcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1198
1112 aaatgctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1171
1199 acttgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1255
1172 agcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1231
1256 aacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1315
1232 tgaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1291
1316 tcaacttgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1375
1292 ggtcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1351
1376 acttgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1435
1352 gaccccaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1382
1436 atccacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1466
RESULT 3
LOCUS 166494/c 7218 bp DNA linear PAT 28-DEC-1997
DEFINITION Sequence 14 from patent US 5670367.
ACCESSION 166494
VERSION 166494.1 GI:2724471
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 7218)
AUTHORS Donner, F., Scheifflinger, F. and Falkner, F. Gunter.
TITLE Recombinant fowlpox virus
JOURNAL Patent: US 5670367-A 14 23-SEP-1997;
FEATURES location/Qualifiers

LOCUS AB045319 1938 bp mRNA linear INV 09-JAN-2002
DEFINITION Chirospalmus quadrigatus mRNA for CqTx-A, complete cds.
ACCESSION AB045319
VERSION AB045319.1 GI:18146992
KEYWORDS
SOURCE ORGANISM
Chirospalmus quadrigatus cDNA to mRNA.
Chirospalmus quadrigatus
Eukaryota; Metazoa; Cnidaria; Cubozoa; Cubomedusae; Chirodropidae;
Chirospalmus.
1 (sites)
Nagai, H., Takuwa-Kuroda, K., Nakao, M., Oshiro, N., Iwanaga, S. and
Nakajima, T.
A Novel Protein Toxin from the Deadly Box Jellyfish (See Wasp,
Habu-Kurage) Chirospalmus quadrigatus
Biosci. Biotechnol. Biochem. 66, 97-102 (2002)
2 (bases 1 to 1938)
Nagai, H.
Direct Submission
Submitted (29-JUN-2000) Hiroshi Nagai, Tokyo University of
Fisheries, 4-5-7, Konan, Minato-Ku, Tokyo 108-8477, Japan
(E-mail:nagai@tokyo-u-fish.ac.jp, Tel:+81-3-5463-0454,
Fax:+81-3-5463-0398)
FEATURES
Source location/Qualifiers
1. 1938
/organism="Chirospalmus quadrigatus"
/db_xref="taxon:130731"
76..1464
/gene="cqt-a"
76..1464
/gene="cqt-a"
/codon_start=1
/product="CqTx-A"
/protein_id="BAB82520.1"
/db_xref="GI:18146993"
translation="MANLYRSLALLFMPTIASCTISSGLASIKAIIDKRRSGCK
LFDVYANNOKIIEKFSNDDERAKWGIISLSTRVGFQSGDPKAINSGCLDILVIG

[illegible]

RESULT	5			
AX251552/c				
LOCUS	AX251552	56153 bp	DNA	linear
DEFINITION	Sequence 520 from Patent WO0168912.			
ACCESSION	AX251552			
VERSION	AX251552.1			
KEYWORDS	GI:15984975			
SOURCE				
ORGANISM	synthetic construct.			
	synthetic construct			
	artificial sequence.			
	1 (bases 1 to 56153)			
REFERENCE	Olek,A., Piepenbrock,C. and Berlin,K.			
AUTHORS	Diagnosis of diseases associated with tumor suppressor genes and			
TITLE	oncogenes			
	Patent: WO 0168912-A 520 20-SEP-2001;			
JOURNAL	Epigenomics AG (DE)			
	Location/Qualifiers			
FEATURES	1..56153			
source	/organism="synthetic construct"			
	/db_xref="taxon:32630"			
	/note="chemically treated genomic DNA (Homo sapiens)"			
BASE COUNT	15738 a 1002 c 12377 g 27036 t			
ORIGIN				

[illegible]

Query Match	2.7%	Score 43.2	DB 6	Length 9733
Best Local Similarity	51.7%	Pred. No. 6.7	Mismatches 113	Indels 2
Matches 123	Conservative 0	Mismatches 113	Indels 2	Gaps 1
QY 1372	ataacataa	atlaaagga	atcaacaat	gltcccaag
Db 9374	AAAAAAT	AACTTAA	CAACAAAT	CTCTAAAA
QY 1432	cgaatgcg	ctactlaa	agtgcaac	ctgtatltt
Db 9314	TTCTACT	ATATATA	ATAATTA	AACTATTA
QY 1492	ataaacca	atccagcg	agactaatt	ccaataa
Db 9254	AAATAAT	ATTAATTA	CAAAATTT	TTTATAT
QY 1552	catgaatc	ctctattt	gacacttt	ccaagagat
Db 9196	TAAATTA	ATATCAAT	TTCTATCT	ATTAATAA

RESULT	8								
AX083744/c									
LOCUS	AX083744	1141 bp	DNA	linear	PAT 28-FEB-2001				
DEFINITION	Sequence 22 from Patent WO0111061.								
ACCESSION	AX083744								
VERSION	AX083744.1	GI:13185472							
KEYWORDS									
SOURCE									
ORGANISM	synthetic construct.								
REFERENCE	synthetic construct								
AUTHORS	artificial sequence.								
TITLE	1 (bases 1 to 1141)								
JOURNAL	Kunst/L. and Clemens,S.								
	Regulation of embryonic transcription in plants								
	Patent: WO 0111061-A 22 15-FEB-2001.								
	UNIVERSITY OF BRITISH COLUMBIA (CA)								
FEATURES	Location/Qualifiers								
source	1..1141								
	/organism="synthetic construct"								
	/db_xref="taxon:32630"								
promoter	1..1141								
	/note="consensus sequence of A.t., L.a., and B.n. PAB1								
	promoters"								
BASE COUNT	123 a	32 c	42 g	112 t	832 others				
ORIGIN									
Query Match	2.7%	Score 42.8	DB 6	Length 1141					

[illegible]

VERSION	KEYWORDS
AP001674.1	GI:7768666
SOURCE	ORGANISM
Homo sapiens DNA.	Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.	
REFERENCE	AUTHORS
1 (sites)	Hattori,M., Fujiyama,A., Taylor,T.D., Matanabe,H., Yada,T., Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Soeda,E., Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K., Poley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R., Patterson,D., Reichwald,K., Rump,A., Schillnabel,M., and Schudy,A. The DNA sequencing of human chromosome 21. The chromosome 21 mapping and sequencing consortium Nature 405 (6784), 311-319 (2000)
TITLE	JOURNAL
2 (passes 1 to 340000)	20289799
REFERENCE	AUTHORS
2 (passes 1 to 340000)	Hattori,M., Fujiyama,A., Taylor,T.D., Matanabe,H., Yada,T., Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Soeda,E., Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K., Poley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R., Patterson,D., Reichwald,K., Rump,A., Schillnabel,M., Schudy,A., Zimmermann,W., Rosenthal,A., Kidoh,J., Shibuya,K., Kawasaki,K., Asakawa,S., Shintani,A., Sasaki,T., Nagane,K., Matsuyama,S., Antonarakis,S.E., Minoshima,S., Shimizu,N., Nordstiek,G., Hornischer,K., Barandt,P., Scharte,M., Schoen,O., Desario,A., Reichelt,J., Kauer,G., Bloeker,H., Ramses,D., Beck,A., Klages,S., Hennig,S., Rieseemann,L., Dagand,E., Wehrmeyer,S., Bortym,K., Gardiner,K., Nizetic,D., Francis,F., Lehrach,H., Reinhardt,R. and Vasp,G.M.L
TITLE	JOURNAL
Direct Submission	Submitted (10-Apr-2000) to the DDBJ/EMBL/GenBank databases. The chromosome 21 Mapping and Sequencing Consortium. * RIKEN Genomic Sciences Center, Human Genome Research Group * Institute of Molecular Biotechnology, Genome Analysis * Keio University School of Medicine, Dept. of Molecular Biology * GBF, Dept. of Genome Analysis * Max-Planck Institute for Molecular genetics (addresses see below)
COMMENT	On May 30, 2000 this sequence version replaced gi:7717271. The chromosome 21 mapping and sequencing consortium consisting of * RIKEN Genomic Sciences Center, Human Genome Research Group, * Sagamihara 228-8555, Japan, * e.mail: hattori@gscl.riken.go.jp/ * URL: http://hgp.gscl.riken.go.jp/ and * Institute of Molecular Biotechnology, Genome Analysis, * Beutenbergstrasse 11, D-07745 Jena, Germany, * e.mail: gscj-submit@genome.imb-jena.de * URL: http://genome.imb-jena.de/ and * Keio University School of Medicine, Molecular Biology, * Tokyo 160-8582, Japan, * e.mail: nshimizu@db.med.keio.ac.jp * URL: http://www.db.med.keio.ac.jp/ and * GBF, Dept. of Genome Analysis, * Mascheroder Weg 1, D-38124 Braunschweig, Germany, * e.mail: info.genome@gbf.de * URL: http://genome.gbf.de/ and * Max-Planck Institute for Molecular Genetics, * Ihnestrasse 73, D-14195 Berlin, Germany, * e.mail: info-ch21@molgen.mpg.de * URL: http://chr21.tz-berlin.mpg.de/ AL165219. Submitted (10-Apr-2000). Location/Qualifiers
FEATURES	SOURCE
1.340000	/organism="Homo sapiens"
	/db_xref="taxon:9606"
<1.6149	/chromosome="21"
	/note="Accession NO. AP001537"
	/organism="Homo sapiens"

```
/db_xref="taxon:9606"
/chromosome="21"
/map="21q21.1"
/clone="R749F18, 5' partial"
/clone_lib="RPC1-11 BAC library"
<!.165231
/note="Accession No. AP001538"
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q21.1"
/clone="R853K11, 5' partial"
/clone_lib="RPC1-11 BAC library"
15..444
/note="LIM4A"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
555..3531
/note="LIM4A"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
3770..3876
/note="L2"
/rpt_family="LINE/L2"
/rpt_type=DISPERSED
complement(5645..6178)
/note="LIM2"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
6181..6534
/note="HE1B"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
8021..8370
/note="HE1B"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
complement(8373..10032)
/note="LIM2"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
10213..10520
/note="AluDb"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
15810..15851
/note="(CA)n"
/rpt_family="Simple_repeat"
/rpt_type= tandem
complement(16628..16697)
/note="MER53"
/rpt_family="DNA"
/rpt_type=DISPERSED
16638..16715
/note="MER53"
/rpt_family="DNA"
/rpt_type=DISPERSED
18847..19585
/note="L2"
/rpt_family="LINE/L2"
/rpt_type=DISPERSED
19795..19917
/note="(TA)n"
/rpt_family="Simple_repeat"
20023..20166
/note="(TA)n"
/rpt_family="Simple_repeat"
20289..20448
/note="(TA)n"
/rpt_type= tandem
/rpt_family="Simple_repeat"
/rpt_type= tandem
```

```
repeat_region 20467..20542
/note="(TA)n"
/rpt_family="Simple_repeat"
/rpt_type= tandem
complement(21360..22028)
repeat_region /note="LIM3"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
complement(22142..22448)
repeat_region /note="AluX"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
complement(22519..22891)
repeat_region /note="MER52A"
/rpt_family="LTR/MER4-group"
/rpt_type=DISPERSED
22976..23239
/note="LIM3"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
complement(23244..23519)
repeat_region /note="MER4C"
/rpt_family="LTR/MER4-group"
/rpt_type=DISPERSED
complement(23644..23861)
repeat_region /note="MER4C"
/rpt_family="LTR/MER4-group"
/rpt_type=DISPERSED
23925..25091
/note="MER52A"
/rpt_family="LTR/MER4-group"
/rpt_type=DISPERSED
complement(25227..25672)
repeat_region /note="MER4A2"
/rpt_family="LTR/MER4-group"
/rpt_type=DISPERSED
complement(25682..25987)
repeat_region /note="AluY"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
26744..26849
/note="MER46B"
/rpt_family="DNA/MER2-type"
/rpt_type=DISPERSED
complement(26852..27629)
repeat_region /note="LTR17"
/rpt_family="LTR/Retroviral"
/rpt_type=DISPERSED
complement(27630..27921)
repeat_region /note="HERV17"
/rpt_family="LTR/Retroviral"

Query Match 2.7%: Score 42.8; DB 9; Length 340000;
Best Local Similarity 47.1%; Pred. No. 6.4;
Matches 131; Conservative 0; Mismatches 147; Indels 0; Gaps 0;

OY 1333 agcttggaataatccagagacctgaaggacattggagtatacataataaaggaggaat 1392
DB 59794 AGAGGAGGAGTAAGTACAGAGATCTATTGTGCCAACATGATACATATGTTATGAAAT 59735
OY 1393 caacaatgcccaaggacatacgatatagaacatacgaatgcagtaactaaagtgc 1452
DB 59734 GATTTGTGATTTTAAGATGCTCATAGTAGTCCTCAATGTCTTCATTACAAAATTA 59675
OY 1453 aaacttgattctacataagattgcgtcatgaagtcataaaccatccagcggaactaa 1512
DB 59674 ATGATAGTTGTGACGTACTAATAATATGCTAAGCTTGATTCACCATCAATAATTA 59615
OY 1513 ttcatataaacaataatgttctctataagaactttcaagaacatcctattgtgaca 1572
DB 59614 CATGTATCAAAACATACATATTATGTATATATCACTATATAAATTTTCCTTTTCA 59555
OY 1573 ttcaagagatgtttgaagaacaacaaaaaaa 1610
```

```

DB 59554 ATTTAAAAATATATTTAAAAAGTATACCAAGAAA 59517
RESULT 15
AC027059/C
LOCUS
DEFINITION Homo sapiens chromosome 4 clone RP11-595J23 map 4, WORKING DRAFT
ACCESSION AC027059
VERSION AC027059.2 GI:7596861
KEYWORDS HTG; HTGS-PHASE1; HTGS-DRAFT.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE 1 (bases 1 to 195721)
JOURNAL Homo sapiens chromosome 4, clone RP11-595J23
AUTHORS Unpublished
2 (bases 1 to 195721)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Bouckhagalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Chongel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Dekrelano,K., Dewar,K., Diaz,J.S.,
Dodgson,S., Domingo,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardy,J., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Larocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Margulis,N.,
McCarthy,M., McEwan,P., McGuck,A., McKernan,K., McPheters,R.,
Murphy,T., Naylor,J., Norman,C.H., Miranda,C., Mleagda,V., Morrow,J.,
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Teafaye,S., Theodore,J., Tyrrell,A., Travers,M., Triggillo,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (26-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 19, 2000 this sequence version replaced gi:7329419.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L7408
Center clone name: 599_J23
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 184383 bases at least Q40
Consensus quality: 193091 bases at least Q30
Insert size: 170000; agarose-fp
Insert size: 194121; sum-of-contigs
Quality coverage: 4.6 in Q20 bases; agarose-fp
Quality coverage: 4.0 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 17 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as

```

```

* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
1652: contig of 1652 bp in length
1653 1752: gap of 100 bp
1753 6562: contig of 4810 bp in length
6563 6662: gap of 100 bp
6663 10464: contig of 3802 bp in length
10465 10564: gap of 100 bp
10565 15408: contig of 4844 bp in length
15409 15508: gap of 100 bp
15509 21412: contig of 5904 bp in length
21413 21512: gap of 100 bp
21513 27609: contig of 6097 bp in length
27610 27709: gap of 100 bp
27710 32894: contig of 5185 bp in length
32895 32994: gap of 100 bp
32995 37528: contig of 4534 bp in length
37529 37628: gap of 100 bp
37629 45785: contig of 8157 bp in length
45786 45885: gap of 100 bp
45886 58042: contig of 12157 bp in length
58043 58142: gap of 100 bp
58143 73270: contig of 15127 bp in length
73270 73369: gap of 100 bp
73370 89506: contig of 16137 bp in length
89507 89606: gap of 100 bp
89607 104179: contig of 14573 bp in length
104180 104279: gap of 100 bp
104280 122927: contig of 18648 bp in length
122928 123027: gap of 100 bp
123028 144803: contig of 21776 bp in length
144804 144903: gap of 100 bp
144904 169867: contig of 24964 bp in length
169868 169967: gap of 100 bp
169968 195721: contig of 25754 bp in length.
Location/Qualifiers
1. 195721
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/map="4"
/clone="RP11-595J23"
/clone_id="RP11-595J23"
1. 1652
/note="assembly-fragment"
1753. 6562
/note="assembly-fragment"
6663. 10464
/note="assembly-fragment"
10565. 15408
/note="assembly-fragment"
15509. 21412
/note="assembly-fragment"
21513. 27609
/note="assembly-fragment"
27710. 32894
/note="assembly-fragment"
32995. 37528
/note="assembly-fragment"
37629. 45785
/note="assembly-fragment"
45886. 58042
/note="assembly-fragment"
58143. 73269
/note="assembly-fragment"
73370. 89506
/note="assembly-fragment"
89607. 104179
/note="assembly-fragment"
clone_end:sp6
vector_side:right"

```


DR WPI: 1999-580740/49.
 DR P-PSDB: AAY33651.
 XX Protein with hemolytic activity, useful for drugs treating jelly fish
 PT stings, pharmaceuticals with blood platelet agglutination activity,
 PT pesticides by use of the hemolytic activity, and study of the hemolytic
 PM mechanism
 PS Disclosure: Page 23-25; 32pp; Japanese.
 XX
 CC This invention describes a novel protein which has hemolytic activity,
 CC blood platelet agglutination activity and a molecular weight of about
 CC 50,000 Da (by SDS-PAGE). The protein, homologs and analogs of the
 CC invention can be used as a novel approach to developing drugs useful for
 CC the treatment of jelly fish stings, pharmaceuticals with blood platelet
 CC agglutination activity, pesticides by use of the hemolytic activity, and
 CC in the study of the hemolytic mechanism. This sequence encodes the
 CC hemolytic protein described in the invention
 CC
 XX Sequence 1610 BP; 509 A; 312 C; 352 G; 437 T; 0 other;
 SO

Query Match 100.0%; Score 1610; DB 20; Length 1610;
 Best Local Similarity 100.0%; Pred. NO. 0;
 Matches 1610; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcacaaagcacttggtgaagagacacacatgattctgaacatctcttgcttatt 60
 DB 1 gcacaaagcacttggtgaagagacacacatgattctgaacatctcttgcttatt 60
 QY 61 gtcccttgcaataactcttcgcaaaacatgcaagcctctgataattcttact 120
 DB 61 gtcccttgcaataactcttcgcaaaacatgcaagcctctgataattcttact 120
 QY 121 aagtgagaactgcctttaaagaagctctctgtagcaagagcctcttgaggctta 180
 DB 121 aagtgagaactgcctttaaagaagctctctgtagcaagagcctcttgaggctta 180
 QY 121 aagtgagaactgcctttaaagaagctctctgtagcaagagcctcttgaggctta 180
 DB 121 aagtgagaactgcctttaaagaagctctctgtagcaagagcctcttgaggctta 180
 QY 181 gaaggcttaaaagagagatccagacaaacagacgagttgagacaaacaaatc 240
 DB 181 gaaggcttaaaagagagatccagacaaacagacgagttgagacaaacaaatc 240
 QY 181 gaaggcttaaaagagagatccagacaaacagacgagttgagacaaacaaatc 240
 DB 181 gaaggcttaaaagagagatccagacaaacagacgagttgagacaaacaaatc 240
 QY 241 ctgtgattctgctgagctcctctctgtagaaatctctgtagatgacacaaatc 300
 DB 241 ctgtgattctgctgagctcctctctgtagaaatctctgtagatgacacaaatc 300
 QY 301 tctgtgtcctcgacatgtgtgcagaaatgcaacaactttggagggcctgtcgagatg 360
 DB 301 tctgtgtcctcgacatgtgtgcagaaatgcaacaactttggagggcctgtcgagatg 360
 QY 361 ggaatcgagagcctgagctcttcttctcctcaattctacatctgcttactggaagctca 420
 DB 361 ggaatcgagagcctgagctcttcttctcctcaattctacatctgcttactggaagctca 420
 QY 421 gaaaggaactcaggtgtgcgcgcttatctatagagctttaaagacatcgatgagcc 480
 DB 421 gaaaggaactcaggtgtgcgcgcttatctatagagctttaaagacatcgatgagcc 480
 QY 481 atccaaagacatgcagcaggtgcagagagagatttctgtaacatcgcgcatcttaag 540
 DB 481 atccaaagacatgcagcaggtgcagagagagatttctgtaacatcgcgcatcttaag 540
 QY 541 gtcatgaaacagcagtcacatctacagatagcagacctaagatcatgacagcagatgt 600
 DB 541 gtcatgaaacagcagtcacatctacagatagcagacctaagatcatgacagcagatgt 600
 QY 601 cctgtttataaattagtaattctatcggaacagttggagagcagaattcccaaggcgca 660
 DB 601 cctgtttataaattagtaattctatcggaacagttggagagcagaattcccaaggcgca 660
 QY 661 gcaactacagctcttagcagatcagaagagagcgttgaacttactctgctcatgttaa 720
 DB 661 gcaactacagctcttagcagatcagaagagagcgttgaacttactctgctcatgttaa 720

QY 721 ctgtgacatgagaagaaccttgctgtgctgactgtgctatctctacagaaagaaat 780
 DB 721 ctgtgacatgagaagaaccttgctgtgctgactgtgctatctctacagaaagaaat 780
 QY 781 gcagaacagctggcagatgctgtggaacagctaataatgggtaaacaaagcagctgct 840
 DB 781 gcagaacagctggcagatgctgtggaacagctaataatgggtaaacaaagcagctgct 840
 QY 841 gataccctaatcttctcctcctaaattgattcctcgtgaagaacagcatgagtgcaattat 900
 DB 841 gataccctaatcttctcctcctaaattgattcctcgtgaagaacagcatgagtgcaattat 900
 QY 901 catcaattctgctccttggaactagcaagaacatataatcaacagaaactcttggga 960
 DB 901 catcaattctgctccttggaactagcaagaacatataatcaacagaaactcttggga 960
 QY 961 gtccagatgtcccgctcctatctggaacgcagatatacaaatcttacaatagttactg 1020
 DB 961 gtccagatgtcccgctcctatctggaacgcagatatacaaatcttacaatagttactg 1020
 QY 1021 aataccctacagatatacagtgagcttaccatgaggaattacatggtcagagctgtct 1080
 DB 1021 aataccctacagatatacagtgagcttaccatgaggaattacatggtcagagctgtct 1080
 QY 1081 aacgttcggaatcccaatatacagtgagtaaccaaaatgctcgtatggtttcacacatgag 1140
 DB 1081 aacgttcggaatcccaatatacagtgagtaaccaaaatgctcgtatggtttcacacatgag 1140
 QY 1141 aatagcagatcggaggaagtgtatatacacaagatgacaaagatggtggatggttact 1200
 DB 1141 aatagcagatcggaggaagtgtatatacacaagatgacaaagatggtggatggttact 1200
 QY 1201 ttgagatgagatccaggttgacaaagcagcatatgagttccttcttgaagactgggaag 1260
 DB 1201 ttgagatgagatccaggttgacaaagcagcatatgagttccttcttgaagactgggaag 1260
 QY 1261 tataatgtaagctctcagaaggtggcccaactggttcatatgatatgataagtgccagt 1320
 DB 1261 tataatgtaagctctcagaaggtggcccaactggttcatatgatatgataagtgccagt 1320
 QY 1321 ggtcaacttcgacgtgtggaaataatccagagacctaagaagatctggaatataacata 1380
 DB 1321 ggtcaacttcgacgtgtggaaataatccagagacctaagaagatctggaatataacata 1380
 QY 1381 ttaagaggaatcaacaatctcccaagagcatcgaatatagaacatcaaacgaatgcag 1440
 DB 1381 ttaagaggaatcaacaatctcccaagagcatcgaatatagaacatcaaacgaatgcag 1440
 QY 1441 tacttaagtgacacatgttatctacatagagatgctgcgatgaaagtcataaaccat 1500
 DB 1441 tacttaagtgacacatgttatctacatagagatgctgcgatgaaagtcataaaccat 1500
 QY 1501 ccaggggactcaattcatatataaactaatgttccctataatgcatcttcaagaatc 1560
 DB 1501 ccaggggactcaattcatatataaactaatgttccctataatgcatcttcaagaatc 1560
 QY 1561 tctatttgacatttcaagagagatgtttgaaagaaacaaaaaaat 1610
 DB 1561 tctatttgacatttcaagagagatgtttgaaagaaacaaaaaaat 1610

RESULT 2
 AA166621
 ID AA166621 standard; DNM; 2042 BP.
 XX
 AC AA166621:
 XX
 DT 07-JAN-2002 (first entry)
 XX
 DE C. alata hemolytic protein encoding DNM.
 XX Hemolytic protein; nematocyst; sting; drug; blood platelet;
 KW

KM	agrochemical; ds.
XX	
OS	Carybdea alata.
XX	
FH	Key
FT	Location/Qualifiers
FT	73..1464
XX	/+tag- a
PN	WO200170799-AI.
XX	
PD	27-SEP-2001.
XX	
PE	21-MAR-2001; 2001WO-JP02209.
XX	
PR	21-MAR-2000; 2000JP-0078967.
PA	(SUNR) SUNTORY LTD.
XX	
PI	Nagai H, Kuroda K, Nakajima T;
DR	WPI; 2001-611481/70.
DR	p-PSDB; AAG65633.
XX	
PT	New polypeptide for treating stings, and for the development of
XX	pharmaceuticals and agrochemicals, and studying hemolysis, comprises
XX	the Carybdea alata-originated protein -
XX	
PS	Examples; Page 23-24; 28pp; Japanese.
XX	
CC	The invention provides a polypeptide from nematocyst of C. alata. The
CC	protein comprises hemolytic activity, and has a molecular weight of about
CC	50,000 Daltons as determined by SDS-Polyacrylamide gel electrophoresis
CC	(SDS PAGE). The protein is useful for treating stings, and also in the
CC	development of drugs with blood platelet aggregating effects, for the
CC	development of agrochemicals with hemolytic effects' and as reagents for
CC	studying hemolytic mechanisms. The present sequence represents the
CC	DNA encoding a C. alata protein having hemolytic activity.
XX	
QX	Sequence 2042 BP; 616 A; 400 C; 424 G; 602 T; 0 other;

[illegible]

OY	641	gcgaatctcccaagggcgacgaacactacacagctcttgagcatgcaagaagagcccttgact	700
Db	719	gcgaatctcccaagggcgacgaacactacacagctcttgagcatgcaagaagagcccttgact	778
OY	701	tcaattcgtctatctgtcaactctgttagtcataagaaacacctgtgcgtgcagcttgcta	760
Db	779	tcaattcgtctatctgtcaactctgttagtcataagaaacacctgtgtactcaagttatctc	838
OY	761	tctctacaggaagaagaatgacgaacag-----tgccaagtgcctgtgaaacgcta	814
Db	839	tctgttacaagcgtgcggtgtgtgcatactatgagcttgcgaactgtctcgtctcttaaca	898
OY	815	ataaggtaaacaagaagcgtcagctgcgtataccctagatcttctcttaaatcttgatccg	874
Db	899	gtgatccaataacaaagaagcgcaagaagaaacggttcaccttcttaccataatggaaca	958
OY	875	aacaagcatgtatagtgtagcagttatcatccaattcttcgtctctgaaactagaacaagca	934
Db	959	agatctctctcgtgtgtcttactactacacctctatgacacactaaggcagcatgtgta	1011
OY	935	tatttaaatcacgaataactcttgaggttcccaatgttcccg---tctatttgaaac	991
Db	1019	tcttcaactcaacaactcttcttgaggtgcgaatccctgcgaatacacagtttgatgctc	1077
OY	992	gcgaatacaaatcttacaataagattgatactgtagataccctagatatgcaatgaggttaca	1051
Db	1079	ttatttacaagatgcaaaaacagggacatgagatcgtgatatgcatctgttaagaactcttatg	1138
OY	1052	tgggaattacatcagttctcaagagcctgtcttaagcttcggaatccaatcagaagtatcca	1111
Db	1139	cgggacatcatcactgttcttcgggtgcgtgcgaagaatccaagtatcacaatgaaatagatcca	1198
OY	1112	aatgtctgcatgggttttacaacatggaagatagcgatcggaggaagtgtgtatatcaca	1177
Db	1199	agcgggaataatggttaccatacctatctaccctcgatccaagaagcca---tgtatgcagca	1255
OY	1172	agcatgacccaagagatggggatggtgtaacttttgagatgaagatccgaagtgcacaagccata	1231
Db	1256	aacacgcctcaagagatgggggtcgtgggaacagatgataagatccaggtggaagagttact	1311
OY	1232	tgaagttcatctcccttgaagacatgggaagatataatgtaagctctaagaagtggcccaact	1291
Db	1316	tcaacttcatccctttaaacaatggttttaccatggttcttaccagaagatggccagatt	1377
OY	1292	ggtcattgatatatgaaatcgaagtgcacagtgcatacttcgaagcttggaataatccag	1351
Db	1376	acttgtgttaactatgaaagcagatgctgcatgtgtatatctcgaagctgtgattacaacccg	1433
OY	1352	gaacctcaaggaacatctggagtatacataat	1382
Db	1436	atccacagggagacatggaataatctgttaatt	1466
RESULT 3			
AAS46794/c			
ID	AAS46794 standard; DNA; 56153 BP.		
XX	AAS46794;		
XX	18-DEC-2001 (first entry)		
XX	Tumour suppressor gene derived chemically modified sequence #520.		
KW	Human; tumour suppressor gene; oncogene; antitumour; cytosolic;		
KW	cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;		
XX	cytosine methylation; ds.		
OS	Homo sapiens.		
XX	WO200168912-A2.		
XX	20-SEP-2001.		

PF	15-MAR-2001; 2001IWO-EPO2955.
PR	15-MAR-2000; 2000DE-1013847.
PR	06-APR-2000; 2000DE-1019058.
PR	07-APR-2000; 2000DE-1019178.
PR	30-JUN-2000; 2000DE-1032529.
PR	01-SEP-2000; 2000DE-1043826.
XX	(EPIG-) EPIGENOMICS AG.
PA	
XX	
PI	Olek A. Piepenbrock C, Berlin K;
XX	
DR	WPI; 2001-602752/68.
XX	
PT	Fragments of chemically modified genes associated with tumour suppressor
PT	genes and oncogenes, useful in designing primers and probes for
PT	analysing diseases associated with cytosine methylation state e.g.
PT	cancer
XX	
PS	Claim 1; SEQ ID No 520; 27pp; English.
XX	
CC	The invention relates to a nucleic acid comprising a sequence of 18
CC	bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC	disulphite, or genes associated with tumour suppression and
CC	oncogenes having a sequence taken from 536 (actually 533 since
CC	numbers 408, 458 and 500 are missing from the sequence listing) sequences
CC	(Ss) and sequences complementary to (Ss). The nucleic acid may be a
CC	peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
CC	form part of a set of probes for detecting the cytosine methylation state
CC	and/or single nucleotide polymorphisms and also to be used in an
CC	array for analysing diseases associated with CpG dinucleotides e.g.
CC	cancers and tumours. The probes can also be used in a method for
CC	ascertaining genetic and/or epigenetic parameters for the diagnosis
CC	and/or therapy of existing diseases or the predisposition to specific
CC	diseases, by analysing cytosine methylations. The parameters may be
CC	compared to another set of genetic and/or epigenetic parameters, the
CC	differences serving as basis for diagnosis and/or prognosis events which
CC	are disadvantageous to patients. The present sequence is one of the
CC	533 genomic sequences derived from tumour suppressor genes and
CC	oncogenes. Sequences with even numbered Seq ID numbers are the
CC	complementary sequence of the corresponding odd numbered sequence (e.g.
CC	ID 2 and ID1, ID 536 and ID 535, except for those whose partner sequence
CC	is missing).
CC	Note: The sequence data for this patent did not form part
CC	of the printed specification, but was obtained in electronic
CC	format directly from WIPO at
CC	ftp.wipo.int/pub/published_pct_sequences.
XX	
SQ	Sequence 56153 BP; 15738 A; 1002 C; 12377 G; 27036 T; 0 Other;
<hr/>	
Query Match	2.9%; Score 46.2; DB 22; Length 56153;
Best Local Similarity	52.2%; Pred. NO. 0.073;
Matches 128; Conservative	0; Mismatches 113; Indels 4; Gaps 1
Dy	1362 acattggagataacataactaaagaaggaatacaaatgcccagaagcatacgaatata 1421
Dy	
Dy	36239 AAATTTCGCAATACAACTTTTATAAACACCACAATCAATTAATTAATAAAAAAATA 36180
Dy	1422 agacataacaagaatcgacttaacttaagaagcacacttgatcttcacataagatgcgc 1481
Dy	
Dy	36179 AACCCAAAACCCCTTAACGTCTACTANA-----ATCCTTTTATTAATATATATATTCCT 36124
Dy	1482 atgaaagtcataaaccalcagcgagactaatcttcatataaacaataatgtttccctat 1541
Dy	
Dy	36123 ATATATCTCCCAAAAACATAAAAAATCTTTTCACCATATATATAACAAATCTCACATCTT 36064
Dy	1542 aatgatcttcataagaatctcatatgtagcatttcagaagggtatgttgaagaanaaa 1601
Dy	
Dy	36063 TATTAATTCACATAAATATAAAAACTTAACACTTATATATATATATTAATAACCTAAAGC 36004
Dy	
Dy	1602 aaaaa 1606

Dd	36003	CATTA	35999
----	-------	-------	-------


```

RESULT      4
ID          ABLJ34155/c
AC          ABLJ34155 standard; DNA: 15548 BP.
XX
XX          ABLJ34155;
XX
XX          26-MAR-2002 (first entry)
XX
Dd          Human immune system associated gene SEQ ID NO: 2128.
Dx
Dx          Human; immune system disease; cytosine methylation; antiasthmatic;
Kw          antidiabetic; anti-HIV; anticancer; cytostatic; neuroprotection;
Kw          neuroprotective; anti-HIV; anticancer; cytostatic; neuroprotection;
Kw          antirheumatic; arthritis; antidiabetic; antiparasitic;
Kw          antinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
Kw          acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
Kw          neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
Kw          gene; ds.
Xx
Xx          Homo sapiens.
Xs
Xs          WO200200928-A2.
Pn
Pn          03-JAN-2002.
Pd
Pd          02-JUL-2001; 2001WO-EP07537.
Pf
Pf          30-JUN-2000; 2000DE-1032529.
Pr          01-SEP-2000; 2000DE-1043826.
Pa          (EPIG-) EPIGENOMICS AG.
Xx
Xx          Olek A. Piepenbrock C, Berlin K;
Pi          WPI; 2002-130909/17.
Dr
Dr          Nucleic acid comprising fragment of chemically modified gene, useful
Pt          for diagnosis and treatment of diseases associated with abnormal
Pt          cytosine methylation -
Ps          Claim 1; SEQ ID NO 2128; 32pp + Sequence Listing; German.
Cc
Cc          The present invention provides a number of human immune system associated
Cc          genes which are modified by the methylation of cytosines. The sequences
Cc          can be used in the diagnosis and treatment of immune system disorders,
Cc          including eye diseases such as retinopathy, neovascular glaucoma and
Cc          macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
Cc          leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
Cc          rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
Cc          diseases. The present sequence is a gene of the invention.
Cc
Sd          Sequence 15548 BP; 4209 A; 247 C; 2903 G; 8189 T; 0 other;

```

Query Match 2.7%; Score 43.8; DB 24; Length 15548;
Best Local Similarity 47.6%; Pred. NO. 0.19;
Matches 129; Conservative 0; Mismatches 142; Indels 0; Gaps 0;

Oy	1340	aaataacccaggagcctcaagaagcattgaggtatatacaataaaggaatcaacaat	1399
Dd	11887	AAAAATAAATAATATTAAATAATTTAAAAAATTTAAAAATTTAAAAATTTAAATAT	11828
Oy	1400	gtcccctaagagcatagaataataagaatcaagaagtgcagtaactaagtgcacctg	1459
Dd	11827	AAAAATTAATAATTTAAAAATTTAAAAATTTAAAAATTTAAAAATTTAAAAATTA	11768
Oy	1460	tatttcataagtagtgcgtcgatcgaagtccataaacaccatccagcgactaatlcata	1519
Dd	11767	AATATTAATAAATAATTAATTAATTAATAATTAATAATTAATAATTAATAATTAATA	11708

OY	1579	ttaaaccataatgctccctaataagcatttcacgaaatccacttgtagcatttaag
Dy	11707	TAAATAATTAATAAATAATAATTATAATTTAAAAATAATAATATAATAATAATAA
OY	1580	aggatatcgttggaaagaacaacaaaaaaaaa 1610
Db	11647	ATAAAATATATAAATAATAATAATAATAATAA 11617
RESULT	5	
ABL32682/c		
ID	ABL32682 standard; DNA; 9733 BP.	
XX	ABL32682:	
AC		
XX		
DT	26-MAR-2002 (first entry)	
XX		
DE	Human immune system associated gene SEQ ID NO: 655.	
XX		
KW	Human: immune system disease; cytosine methylation; antiasthmatic;	
KW	antiarteriosclerotic; antianemic; cytostatic; noctologic;	
KW	neuroprotective; anti-HIV; anticoagulant; ophthalmological;	
KW	antirheumatic; antiarthritis; antidabetic; antipsoriasis;	
KW	antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;	
KW	acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;	
KW	neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;	
KM	gene; ds.	
OS	Homo sapiens.	
PN	NO200200928-A2.	
PN	NO200200928-A2.	
XX		
PD	03-JAN-2002.	
XX		
PF	02-JUL-2001; 2001MO-EP07537.	
XX		
PR	30-JUN-2000; 2000DE-1032529.	
PR	01-SEP-2000; 2000DE-1043826.	
XX	(EPIC-) EPIGENOMICS AG.	
PA		
DR	Olek A, Piepenbrock C, Berlin K;	
PI	WPI; 2002-130909/17.	
Nucleic acid comprising fragment of chemically modified gene, useful for diagnosis and treatment of diseases associated with abnormal cytosine methylation -		
Claim 1; SEQ ID NO 655; 32pp + Sequence Listing; German.		
The present invention provides a number of human immune system associated genes which are modified by the methylation of cytosines. The sequences can be used in the diagnosis and treatment of immune system disorders, including eye diseases such as retinopathy, neovascular glaucoma and macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis, rheumatoid arthritis, psoriasis and inflammatory/allergic bowel diseases. The present sequence is a gene of the invention.		
Sequence 9733 BP; 2955 A; 93 C; 1898 G; 4787 T; 0 other;		
Query Match	2.7%; Score 43.2; DB 24; Length 9733;	
Best Local Similarity	51.7%; Pred No 0.22;	
Matches 123; Conservative 0; Mismatches 113; Indels 2; Gaps 1		
OY	1372	atacaataaattaaagaagaataccaactgtcccagaagcatagaataaagaacataaa 1431
Db	9374	AAATAAAGCTAACCAAAACGACAAATAATTCCTTAAAAAAATTATAATAATAA 9315
OY	1432	cgaatgcagctaacaaagcacaaccttatttcacatagagtgcgatgaagtc 1491

D6	9314	TTCACGTAATATAAATAAAAGAAAAACCTATATATAACTATCATCTATAAATCTTAAAATACC	9255
OY	1492	ataaacatccagcgactaatcattcataattaacaattaaigtcttcctaatagcatlctt	1551
D6	9254	--AATAATAATTAAATTACCAATTTTTATTTATACCTTAATAATATACGAATATAAATAATTT	9197
OY	1552	cacgaacatcttatgtgcacctccaagaaggatatgtttgagaagaaaaaaa	1609
		- - - - -	
D6	9196	TAAATATAATCCATTCTTCTATCTATATAATAAACCTACTGTGTAATAAAAAATACATATAAAA	9139
RESULT	6		
ABL32711/c			
ID	ABL32711	standard; DNA; 7302 BP.	
XX	XX		
DT	26-MAR-2002	(first entry)	
DE	Human immune system associated gene SEQ ID NO: 684.		
XX	XX		
KW	Human; immune system disease; cytosine methylation; antiasthmatic;		
KW	antiarteriosclerotic; antianemic; cytostatic; nootropic;		
KW	neuroprotective; anti-HIV; anticoagulant; ophthalmological;		
KW	antirheumatic; antiarthritic; antididiabetic; antipsoriatic;		
KW	antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;		
KW	acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;		
KW	neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;		
KW	gene; ds.		
OS	Homo sapiens.		
SN			
PX	WO200200928-A2.		
PD			
PF	03-JAN-2002.		
PP	02-JUL-2001; 2001WO-EP07537.		
PR	30-JUN-2000; 2000DE-1032529.		
PR	01-SEP-2000; 2000DE-1043826.		
XX			
PA	(EPIG-) EPIGENOMICS AG.		
XX			
PI	Olek A, Piepenbrock C, Berlin K;		
DR	WIJ: 2002-130909/17.		
XX			
Nucleic acid comprising fragment of chemically modified gene, useful PT for diagnosis and treatment of diseases associated with abnormal PT cytosine methylation - XX			
PS	Claim 1; SEQ ID NO 684; 32pp + Sequence Listing; German.		
XX			
CC	The present invention provides a number of human immune system associated genes which are modified by the methylation of cytosines. The sequences can be used in the diagnosis and treatment of immune system disorders,		
CC	including eye diseases such as retinopathy, neurovascular glaucoma and CC macular degeneration, arteriosclerosis, anemia, cancer, acute myeloid leukemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis, CC rheumatoid arthritis, psoriasis and inflammatory/intestinal bowel diseases. The present sequence is a gene of the invention. CC XX		
SQ	Sequence 7302 BP; 1703 A; 147 C; 1814 G; 3638 T; 0 other:		
Query Match	2.7%; Score 42.8; DB 24; Length 7302;		
Best Local Similarity	47.4%; Pred. No. 0.25;		
Matches 128; Conservative 0; Mismatches 142; Indels 0; Gaps 0;			
D6	1340	aataataaccggagccctcaagagacattggagtatacaataaaggaggaatcaccaat	1399
D6	6955	AAAAAAAAACCTTAATAAATAATCTATAAATAATAAATAAATAAATAAATAAATAAATAAATAA	6896

Oy	1400	gtccaaaggcatcagcaatataagacataaacgtagtgaacttaagtacaacttg	1459
Db	6685	CTCCATTAATAAAGTAAATTAAATTAATTAATTAATTTTCACAAAAACAGCATTTGC	6836
Oy	1460	tattctcacataagatgctgcataagaagtcacataaccatccagcggaactatcata	1519
Db	6683	AATTTAAAAAAAAATAAAAATTCGCAAAACGAATATATTAATTAATTTACACAACATATA	6776
Oy	1520	ttaaaccatbaigtcttcctataatgacattlccaagaatctctatctlygacattcaa	1579
Db	6775	AATTAATTTTCATCTACTCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTTATAT	6716
Oy	1580	aggatagtttgaagaagaacaaaaaaa	1609
Db	6715	TATTAATTAATTTTACACAATAATAAACAAAA	6686
RESULT	7		
AS46746/c			
ID	AS46746 standard; DNA:	38342 BP.	
XX	AS46746;		
AC			
XX			
DT	18-DEC-2001 (first entry)		
XX			
DE	Tumour suppressor gene derived chemically modified sequence #470.		
KW	Human; tumour suppressor gene; oncogene; antitumour; cytostatic;		
KM	cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;		
KV	cytosine methylation; ds.		
OS	Homo sapiens.		
PN	WO200168912-A2.		
PD	20-SEP-2001.		
PF	15-MAR-2001; 2001MO-EP02955.		
PX	15-MAR-2000; 2000DE-1013847.		
PR	06-APR-2000; 2000DE-1019058.		
PR	07-APR-2000; 2000DE-1019173.		
PR	30-JUN-2000; 2000DE-1032529.		
PR	01-SEP-2000; 2000DE-1043826.		
PA	(EPIC-) EPIGENOMICS AG.		
XX			
PI	Olek A, Piepenbrock C, Berlin K;		
DR	WPI; 2001-602752/68.		
XX			
PT	Fragments of chemically modified genes associated with tumour suppressor		
PT	genes and oncogenes, useful in designing primers and probes for		
PT	analysing diseases associated with cytosine methylation state e.g.		
PT	cancer		
XX			
PS	Claim 1; SEQ ID No 470; 27pp; English.		
CC	The invention relates to a nucleic acid comprising a sequence of 18		
CC	bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with		
CC	bisulphite, of genes associated with tumour suppression and		
CC	oncogenes having a sequence taken from 536 (actually 533 since		
CC	numbers 408, 458 and 500 are missing from the sequence listing) sequences		
CC	(Ss) and sequences complementary to (Ss). The nucleic acid may be a		
CC	peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may		
CC	form part of a set of probes for detecting the cytosine methylation state		
CC	and/or single nucleotide polymorphisms and also to be used in an		
CC	array for analysing diseases associated with CpG dinucleotides e.g.		
CC	cancers and tumours. The probes can also be used in a method for		
CC	ascertaining genetic and/or epigenetic parameters for the diagnosis		
CC	and/or therapy of existing diseases or the predisposition to specific		
CC	diseases, by analysing cytosine methylations. The parameters may be		
CC	compared to another set of genetic and/or epigenetic parameters, the		

CC differences serving as basis for diagnosis and/or prognosis events which
CC are disadvantageous to patients. The present sequence is one of the
CC 533 genomic sequences derived from tumour suppressor genes and
CC oncogenes. Sequences with even numbered Seq ID numbers are the
CC complementary sequence of the corresponding odd numbered sequence (e.g.
CC ID 2 and ID1, ID 536 and ID 535, except for those whose partner sequence
CC is missing).

CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 38342 BP; 12471 A; 472 C; 7153 G; 18246 T; 0 other;

Query Match 2.6%; Score 41.2; DB 22; Length 38342;
Best Local Similarity 48.3%; Pred.No.1.6;
Matches 115; Conservative 0; Mismatches 123; Indels 0; Gaps 0

Gy 1372 ataacaataatcaagaaggaatcaacaatggtcccaaggacatcgaatatagaacatcaaa 1431
Db 17739 AAAAATTTAAATAAACCTTACACATATTAACAACCTMAAAATATATATTACTATMAAAAAAACTAC 17670

Gy 1432 cgaatgcagctacttaagaagtgcacacttgattcttcacatagatgtagcgcatgaagtcc 1491
Db 17669 TTATATACCTTTCAAAAACACTCCTTTACTTAACCTTAATATCATATATATATATAAATAAC 17610

Gy 1492 ataacaccatccagcgagactaatlcatatlcaacatlaatglttccctaataatgatctt 1551
Db 17609 TAATATTTTTCCGCCCAAATAATTTCATATATGCAATTTAAACAGCTTAAATTAATTAA 17550

Gy 1552 catgaatcctctatctgtgacatcttcaagaagagatcgtttgaagaagaaaaaaa 1609
Db 17549 AATTTATTTATATATTTCCAAAACACATCAAAAATACCTTTTATATATTATMAAAATFAAA 17492

RESULT 8
AAI61370
ID AAI61370 standard; DNA; 127197 BP.
XX AAI61370;
AC
XX
DT 16-OCT-2001 (first entry)
DE Soybean S15002 region G2, SEQ ID NO: 1.
XX
XX
KW Soybean; anthelmintic; gene therapy; soybean cyst nematode; SCN;
RV SCN resistance; rhg1; Rhg4; SCN resistant allele; plant breeding;
KW 240017 region G3; 318013 region A3; S15002 region G2; ds.
XX
OS Glycine max.
XX
XX W0200151627-A2.
PN
XX
PD 19-JUL-2001.
XX
XX 05-JAN-2001; 2001WO-US00552.
PF
XX 07-JAN-2000; 2000US-0174880.
PR
XX (MONS) MONSANTO CO.
PA
XX Hauge BM, Wang ML, Parsons JD, Parnell LD;
PI
XX WPI: 2001-425872/45.
DR
XX New purified nucleic acid for producing a soybean plant having soybean
PT cyst nematode resistance and for use in plant breeding programs -
XX
PS Disclosure; Page 131-204; 1353pp; English.

The invention relates to nucleic acid molecules from regions of the soybean genome which are associated with soybean cyst nematode (SCN)

CC resistance. The nucleic acids are used to transform plants, and can
CC produce soybean plants having an Rhg1 or an Rhg4 SCN resistant allele.
CC The nucleic acids can be used for investigating rhg1 or Rhg4 haplotypes
CC of soybean plants and for introgressing SCN resistance or partial SCN
CC resistance into soybean plants. They can also be used in plant breeding
CC programmes. The invention also relates to proteins encoded by such
CC nucleic acid molecules, as well as antibodies capable of recognising
CC these proteins. The present sequence is a nucleic acid molecule
CC provided in the specification.

SQ Sequence 127197 BP; 43836 A; 20975 C; 20437 G; 41949 T; 0 other;

	Query Match	Similarity	2.6%	Score 41.2	DB 22	Length 127197	
	Best Local	Similarity	51.8%	Pred.	No. 2.9	Matches 117	Conservative 0; Mismatches 108; Indels 1; Gaps 1
Qy	1372	ataacataatlaaagaagaaatcaacaatgctcccaagagcatcgaataataagacatcaa	1431				
Db	95052	ataacataatlaaataagaacataaataaaaaaataataataaaaaaagaaaaaa	95111				
Qy	1432	cgaatgcagtaactaaagcgacactgttatttctaactagatgctgcgaatgaagctc	1491				
Db	95112	gtctcaatattacagccctaagaatacttaactcttaatbaaggttgggttgaaatc	95171				
Qy	1492	ataaacacatccagcgactaatltatataatlaaacatgaatgttcccttaagacatc	1551				
Db	95172	atlaa-atlaaagtaatttttagaataataaataatacatctttatttaaactaat	95230				
Qy	1552	catgaatctcatatgtgacatttcaagaagatattgtttgaagaa	1597				
Db	95231	atttaatacttcaattgacacatttaataataataatlaaactaattgaagta	95276				

RESULT	9
ABL32347/c	
ID	ABL32347
XY	standard; DNA; 5216 BP.

DT 26-MAR-2002 (first entry)
 XY

Human immune system associated gene SEQ ID NO: 320.

Human; immune system disease; cytosine methylation; antisthmatic
antiartherosclerotic; antanaemic; cyrostatic; nocropic;
neuroprotective; anti-HIV; anticomvulsant; ophthalmologic;
anti-rheumatic; antithyritic; antidiabetic; antipsoriatic;
antiinflammatory; cancer; Eylem disease; arteriosclerosis; anemia;
acute myeloid leukemia; Alzheimer's disease; AIDS; epilepsy;
neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
gene; ds.

Homo sapiens.

PN WO200200928-A2.
xy

PD 03-JAN-2002.
XX

02-JUL-2001; 2001WO-EP07537

PR	01-SEP-2000: 2000DE-1043826
PR	30-JUN-2000: 2000DE-1032529
PR	

PA (EPiG-) EPiGENOMICS AG.

PI Olek A, Piepenbrock C,

WPI; 2002-130909/17.

Nucleic acid comprises

cytosine methylation -

XX
PS
XX

Claim 1; SEQ ID NO 320; 32pp + Sequence Listing; German.

The present invention provides a number of human immune system associated genes which are modified by the methylation of cytosines. The sequences can be used in the diagnosis and treatment of immune system disorders, including eye diseases such as retinopathy, neovascular glaucoma and macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis, rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel diseases. The present sequence is a gene of the invention.

sequence 5216 BP; 1332 A; 73 C; 1196 G; 2615 T; 0 other;

[illegible]

RESULT: 10
ABL34459/c
ID ABL34459 standard; DNA; 5216 BP

AC ABL34459;
VV

DT 26-MAR-2002 (first entry)
 YY

Human metastasis associated gene SEQ ID NO: 12

KW Metastasis associated gene; cytostatic; gene therapy; cancer;
KW cytosine methylation; gene; ds.
KW

US Homo sapiens
XY

PN WO2001/1316-A2.
XX

18-OCJ-2001-
PD
XX

XX 00-AFK-2001; 2001WC-EP039/0

PR 07-APR-2000; 2000DE-1019173.

PR 01-SEP-2000; 2000DE-1043826.

PA (EPIC-) EPIGENOMICS AG.

Olek A, Plepenbrock C, Ber

WPI; 2002-010922/01.

PT New nucleic acid derived from chem

New nucleic acid derived from chemically treated metastasis genes.

DR WPI; 2002-130909/17.

PI Olek A, Piepenbrock C, Berlin K

Query Match	2.2%	Score 34.8;	DB 2;	Length 4165;
Best Local Similarity	51.3%;	Pred. No. 1.9;		
Matches	81;	Conservative	0;	Mismatches 77; Indels 0; Gaps 0;

RESULT 8
US-09-188-930-27
; Sequence 27, Application US/09188930A

```

1.  APPLICANT:  Mullison, James Greg
2.  TITLE OF INVENTION:  Compositions Isolated From Skin Cells
3.  TITLE OF INVENTION:  And Methods For Their use
4.  FILE REFERENCE:  11000.10111
5.  CURRENT APPLICATION NUMBER:  US/09/188,930A
6.  CURRENT FILING DATE:  1998-11-09
7.  NUMBER OF SEQ. ID NOS:  348
8.  SOFTWARE:  FastSeq for Windows Version 3.0
9.  SEQ ID NO 27
10. LENGTH:  778
11. TYPE:  DNA
12. ORGANISM:  Rat
13. US-09-188-930-27

Query Match      2.18;  Score 34;  DB 3;  Length 778;
Best Local Similarity  54.2%;  Pred. No. 1.3;

```

```

RESULT          9
US-09-188-930-202
: Sequence 202, Application US/09188930A
: Patent No. 6150502
: GENERAL INFORMATION:
:   APPLICANT: Watson, James D.
:   APPLICANT: Strachan, Lorna
:   APPLICANT: Sleeman, Matthew
:   APPLICANT: Onrust, Rene
:   APPLICANT: Murison, James Greg
:   TITLE OF INVENTION: Compositions Isolated From Skin Cells
:   TITLE OF INVENTION: and Methods For Their Use
:   FILE REFERENCE: 11000.1011c1
:   CURRENT APPLICATION NUMBER: US/09/188,930A
:   CURRENT FILING DATE: 1998-11-09
:   NUMBER OF SEQ. ID NOS: 348
:   SOFTWARE: FastSeq for Windows Version 3.0
:   SEQ ID NO 202
:   LENGTH: 778
:   TYPE: DNA
:   ORGANISM: Rat
US-09-188-930-202

```

```

Query Match Similarity      2.1%; Score 34; DB 3; Length 778;
Best Local Similarity      54.2%; Pred. No. 1.3;
Matches 90; Conservative 0; Mismatches 75; Indels 1; Gaps 1;

QY      803  tggaaacgctaataatggglaaacaagaagctagctgctgataaccctagattcttcata 862
          |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db       2  tgcagctgcacactagctgataccaagattcgcaacgagataaaggacatttgcctcata 61
          |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY      863  aattatccctgaacaagcatgtgataggtgc-gttatcaccaatttcctcctgaa 921
          |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db       62  aatataaaaaaaggaaattctactactcagccgcactgttcagtcaccacaaatttgaagtaca 121
          |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY      922  actagcaaaagcatatttaattacaacgaataactttggaattccag 967
          |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db       122  aatgaatggaaaacatttatcaccaaatlaattacaattctag 167
          |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

RESULT 10
5352450-1
Patent No. 5352450
APPLICANT: KOGA, TOSHIHIKO; OKHASHI, NOBUO; TAKAHASHI, ICHIRO;
SHIBUYA, KOJI; OHTA, HIROTAKA
TITLE OF INVENTION: METHOD FOR PREPARING VACCINE FOR DENTAL
CARIES AND VACCINAL COMPOSITIONS FOR DENTAL CARIES USED AS NASAL
; DROP
; NUMBER OF SEQUENCES: 2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/529,602
; FILING DATE: 29-MAY-1990
; SEQ ID NO.1:
; LENGTH: 5177
5352450-1

```

RESULT 11
US-08-742-185-101/c
: Sequence 101, Application US/08742185
: Patent No 6020476
: GENERAL INFORMATION:
: APPLICANT: Page, David C.
: APPLICANT: Reljfo, Renee
: APPLICANT: Saxena, Richa
: APPLICANT: Hawkins,Trevor
: APPLICANT: Reeve, Mary Pat
: TITLE OF INVENTION: DAZ: A GENE FAMILY ASSOCIATED WITH AZOOSPERMIA
: NUMBER OF SEQUENCES: 102
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
: STREET: Two Millita Drive
: CITY: Lexington
: STATE: Massachusetts
: COUNTRY: US

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent-in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/742,185
  FILING DATE: 30-OCT-1996
  CLASSIFICATION: 435
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US 08/690,734
    FILING DATE: 31-JUL-1996
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US 08/310,429
    FILING DATE: 22-SEP-1994
  ATTORNEY/AGENT INFORMATION:
    NAME: Granahan, Patricia
    REGISTRATION NUMBER: 32,227
    REFERENCE/DOCKET NUMBER: WH194-07A2
    TELECOMMUNICATION INFORMATION:
      TELEPHONE: (617) 861-6240
      TELEFAX: (617) 861-9540
  INFORMATION FOR SEQ ID NO: 101:
    SEQUENCE CHARACTERISTICS:
      LENGTH: 43795 base pairs
      TYPE: nucleic acid
      STRANDEDNESS: single
      TOPOLOGY: linear
  MOLECULE TYPE: DNA (genomic)
  US-08-742-185-101

```

Query Match 2.1%; Score 34; DB 3; Length 43795;

Query Match	2.1%	Score 33.2	DB 4	Length 1588
Best Local Similarity	45.9%	Pred. No. 3.3		
Matches 113	Conservative 0	Mismatches 133	Indels 0	Gaps 0
QY 420	acgaagaacactcagctgtcgcgccttatcagatgagctttaagcaagacacgcgatgagc	479		
Db 587	AGCAAAAGATTTCATTAATCTAGATTTTGTGAATTACAAATTAAGATTTAGATTATTTCC	646		
QY 480	catccaagacatcgcagcagcagtggtccaaagaagatttgcctgatactcgccttccta	539		
Db 647	CAATGATATCATATATATACAGTGTGAAAAAGAGATTTATTAATATCTCATCAAAATATTC	706		
QY 540	ggtcatggaacagcagcagtcacatcctacagatagcagcactaaagatcatcattgcagcgaatg	599		
Db 707	TGGAACCAATTAATAATCTTAAGACCAATCATATGAAGAATTATTTATTCAGGAAAAAAAAGT	766		
QY 600	tccggttataaattagtaatttatccggagcagcttggagagcagaattcccaagcgc	659		
Db 767	TTTCAATGATATTTTTCAAAAACTGATTATAGTTTGTGAATAATATTAAGACAAAAAANAC	826		
QY 660	agcaac 665			
Db 827	TGTAAAC 832			

Query Match	2.1%	Score 33.2	DB 4	Length 4146
Best Local Similarity	45.9%	Pred. No.5.8	Mismatches 133	Indels 0
Matches 113	Conservative 0	Mismatches 133	Indels 0	Gaps 0
QY	420	agcaagaagcactcagttgctgcgcttatatgatagaagcttgaagcaagcagatcgatgaggc	479	
Db	1942	AGCAAGAAGTTTCAATTACTGATTTTGTGAATTCACAAATAGAGTTTATGATATATTTCCC	2001	
QY	480	catccaagaagacatgcaagcaggtgcccagaagagatttgcctgaatcatctgcattcatca	539	
Db	2002	CAATGTAATTCATTAATACAGTTGTAAGAAAGGAGATTATTTAATATCTCATCAAAATATGCC	2061	
QY	540	ggtcatgaagaagcagctccaactctctacagatagcgaaccatgaatcatctgcagcaatgct	599	
Db	2062	TGGAAACATTAATAATCTCTAAGACCACAAATCAATAGAAAGTTATTTATTCAGGAAAAAAAAGT	2121	
QY	600	tccgtgtataaattatagtaattatctgcagcagttgtaagcgcaaatcttcccagaagcgc	659	
Db	2122	TTTCATGATATATTTTTCACAAAGCTGATATATGTTTGTGAATAATATAGCAACAAAAAAGC	2181	
QY	660	agcaaac	665	
Db	2182	TGTAAC	2187	

LOCUS	LOCUS	LOCUS	LOCUS
CNS03V8B	CNS03V8B	CNS03V8B	CNS03V8B
DEFINITION	Tetraodon nigroviridis genome survey sequence T7 and clone 061H15 of library G from Tetraodon nigroviridis, genomic survey sequence.	Tetraodon nigroviridis genome survey sequence.	Tetraodon nigroviridis genome survey sequence.
ACCESSION	AL262100	AL262100	AL262100
VERSION	GI:7983726	GI:7983726	GI:7983726
KEYWORDS	GSS; genome survey sequence.	GSS; genome survey sequence.	GSS; genome survey sequence.
SOURCE	Tetraodon nigroviridis.	Tetraodon nigroviridis.	Tetraodon nigroviridis.
ORGANISM	Tetraodon nigroviridis	Tetraodon nigroviridis	Tetraodon nigroviridis
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorphi; Tetraodontiformes; Tetraodontidae; Tetraodon.	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorphi; Tetraodontiformes; Tetraodontidae; Tetraodon.	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorphi; Tetraodontiformes; Tetraodontidae; Tetraodon.
AUTHORS	1 (bases 1 to 819) Reest-Ciollius,H., Jalllon,O., Dasilva,C., Fizames,C., Fisher,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.	1 (bases 1 to 819) Reest-Ciollius,H., Jalllon,O., Dasilva,C., Fizames,C., Fisher,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.	1 (bases 1 to 819) Reest-Ciollius,H., Jalllon,O., Dasilva,C., Fizames,C., Fisher,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.
TITLE	Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis	Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis	Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
JOURNAL	Unpublished	Unpublished	Unpublished
REFERENCE	2 (bases 1 to 819)	2 (bases 1 to 819)	2 (bases 1 to 819)
AUTHORS	Bernot-Ciollius,H., Jalllon,O., Dasilva,C., Bouneau,L., Fisher,C., Reest,A., Fizames,C., Wincker,P., Brothier,P., Quetier,F., Saurin,W. and Weissenbach,J.	Bernot-Ciollius,H., Jalllon,O., Dasilva,C., Bouneau,L., Fisher,C., Reest,A., Fizames,C., Wincker,P., Brothier,P., Quetier,F., Saurin,W. and Weissenbach,J.	Bernot-Ciollius,H., Jalllon,O., Dasilva,C., Bouneau,L., Fisher,C., Reest,A., Fizames,C., Wincker,P., Brothier,P., Quetier,F., Saurin,W. and Weissenbach,J.
TITLE	Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence	Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence	Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence
JOURNAL	Unpublished	Unpublished	Unpublished
REFERENCE	3 (bases 1 to 819)	3 (bases 1 to 819)	3 (bases 1 to 819)
AUTHORS	Genoscope.	Genoscope.	Genoscope.
TITLE	Direct Submission	Direct Submission	Direct Submission
JOURNAL	Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases	Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases	Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
COMMENT	This sequence is a single read and was generated as part of a large scale clone-and-sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon .	This sequence is a single read and was generated as part of a large scale clone-and-sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon .	This sequence is a single read and was generated as part of a large scale clone-and-sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon .
FEATURES	location/Qualifiers	location/Qualifiers	location/Qualifiers
Source	1..819	1..819	1..819
	/organism="Tetraodon nigroviridis"	/organism="Tetraodon nigroviridis"	/organism="Tetraodon nigroviridis"

```

BASE COUNT      427 a      13 c      72 g      259 t      48 others
ORIGIN

```

	Query Match	3.1%	Score 49.6	DB 12	Length 819
	Best Local Similarity	47.3%	Pred. No. 0.39		
	Matches 115	Conservative 6	Mismatches 122	Indels 0	Gaps 0
QY	1368	gagttacataatctaaagaggatcaacaatgctcccaaggcagatcgaataagacat	1427		
Db	486	GAATTTAAANANATATANNGAGCAAAAAAATTAATAAAGAAARCGAAGAAAGAA	545		
QY	1428	caaacgaatgcagtaactaaagtcgacacctgtatcttcacatagaatgctgcataa	1487		
Db	546	AAAAAAT	605		
QY	1488	gtccctaaccatccagcgagactatctcatatctaacttaactgcttcctataatgca	1547		
Db	606	ATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	665		
QY	1548	tttccatgaaatctcatatgtgacacttccaaaggataltgtttgaaagaacaaaaaaa	1607		
Db	666	TAAAAAAT	725		
QY	1608	aaa 1610			
Db	726	ATA 728			

RESULT	2
CNS04E5M/C	
LOCUS	976 bp DNA linear GSS 21-MAY-2000
DEFINITION	Tetradodon nigroviridis genome survey sequence T7 end of clone 103P02 of library G from Tetradodon nigroviridis, genomic survey sequence.
ACCESSION	CNS04E5M
VERSION	AL286627
KEYWORDS	GSS; genome survey sequence.
SOURCE	Tetradodon nigroviridis.
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorphi; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetradodon.
REFERENCE	1 (bases 1 to 976)
AUTHORS	Roest-Crollius,H., Jalllon,O., Dasilva,C., Fizesac,C., Fisher,C., Bouneau,L., Billault,A., Queller,F., Saurin,W., Bernot,A. and Weissenbach,J.
TITLE	Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetradodon nigroviridis
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 976)
AUTHORS	Roest-Crollius,H., Jalllon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizesac,C., Winkler,P., Brotlier,P., Queller,F., Saurin,W. and Weissenbach,J.
TITLE	Human gene number estimate provided by genome wide analysis using Tetradodon nigroviridis DNA sequence
JOURNAL	Unpublished
REFERENCE	3 (bases 1 to 976)
AUTHORS	Genoscope.
TITLE	Direct Submision
COMMENT	Submitted (12-Apr-2000) to the EMBL/Genbank/DBJ databases This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetradodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon .
FEATURES	location/Qualifiers
Source	1..976 /organism="Tetradodon nigroviridis"

```

/db_xref="taxon:99883"
/clone="103p02"
/clone_11b="G"
/notes="Genoscope sequence ID : COBG103DH01LPI-end : T7"
BASE COUNT      243 a      151 c      137 g      327 t      118 others
ORIGIN

```

Query Match	3.0%;	Score 49;	DB 12;	Length 976;
Best Local Similarity	38.5%;	Pred. No. 0.54;		
Matches	95;	Conservative	34;	Mismatches 118; Indels 0; Gaps 0;
OY	1364	attgaggtatacacataaattaagaaggaatcaacaatgctcccaagagcatacgaatataag	1423	
Db	914	ATTTTAAAAATTAAAAATTAAAAAAATTTATATTTTNTTAAAAAAMMMWMTAAATTAATAA	855	
OY	1424	acatcaaacgagatgcagctactaaagctgcacactgtatcttcacatagatgctgcacat	1483	
Db	854	ATTTTAAAAAMMMWMTAAATTAATAAATTTATAAAAAAATTTMTNTTTNTTTAAATTTTAATTA	795	
OY	1484	gaaggtccataaacacatccgcgcgcactaatltcatatataaacatataagttccctataa	1543	
Db	794	TAAAAAATAATMMWMTTAAAMTTATNTATNTAMAAAAAATAATWTAATAATWTAATWTA	735	
OY	1544	tgcatttccatgaatctcctatctgtgcacatttcaagagagatagctgttgaagaagaacaa	1603	
Db	734	ATTAAATTATTAATTAATTAATWTAAMNTAAATTTTAAATWMTTNNATTTAAAMATAATAATAA	675	
OY	1604	aaaaaaa 1610		
Db	674	WMHHWA 668		

```

RESULT      3
CNS00396/c
LOCUS
DEFINITION  CNS00396          1101 bp      DNA          linear      GSS 03-JUN-1999
Drosophila melanogaster genome survey sequence TET3 end of BAC #
BACR08K10 of Rpci-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
AL063921
AL063921.1  GI:4941778
GSS.
fruit fly.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1101)
Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : :
BP 191 91006 EVRI cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aron Mammosser in Pletier de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named Rpci-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
p1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila\_bac.htm.
Location/Qualifiers
1..1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="Rpci-98"
FEATURES
SOURCE

```

FEATURES	Location/Qualifiers
source	1 000
pbeloBAC11.	

BASE COUNT	ORIGIN
147 a	111 c 100 g 433 t 51 others

Query Match	2.7%	Score 44.2	DB 12	Length 832
Best Local Similarity	48.2%	Pred. No. 7.1		
Matches 94	Conservative 9	Mismatches 92	Indels 0	Gaps 0

[illegible]

RESULT	6
CNS07EBR/c	
LOCUS	CNS07EBR
DEFINITION	T7 end of clone XBD0AA002B08 of library XBD0AA from strain CBS 94
ACCESSION	A14A1A57
	of <i>Candida tropicalis</i> , genomic survey sequence.
	964 bp DNA linear GSS 08-JUL-2001

VERSION	AL441457.1	GI:12224683
KEYWORDS	GSS.	
SOURCE	Candida tropicalis.	
ORGANISM	Candida tropicalis	

REFERENCE AUTHORS	Saccharomycetales: mitosporic 1 (bases 1 to 964)	Candida.
Blandin, G., Ozier-Kalogeropoulos, O., Wincker, P., Artiguenave, F. and		

TITLE	Genomic exploration of the hemiascomycetous yeasts: 16. <i>Candida tropicalis</i>
JOURNAL	FEBS Lett. 487 (1), 91-94 (2000)
MEDLINE	20584726
REFERENCE	2 (bases 1 to 964)
AUTHORS	Soucieu, J.L., Aigle, M., Artiguenave, F., Blandin, G., et al.

de-Montigny, J., Dujon, B., Durieux, P., Lepingle, A., Llorente, B., Malperru, A., Neuvéglise, C., Olier-Kalogeropoulos, O., Potier, S., Saitou, M., Tekait, F., Toffano-Nioche, C., Wesolowski-Louvel, M., Winkler, P., and Weissenbach, J.

TITLE
Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies
FEBS Lett. 487 (1), 3-12 (2000)

REFERENCE	3 (bases 1 to 964)
AUTHORS	Genoscope.
TITLE	Direct Submission
JOURNAL	Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage, 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail: seque@genoscope.cns.fr Web : http://www.genoscope.cns.fr)
COMMENT	This GSS is part of a random genomic sequencing program of thirteen

yeast species: Saccharomyces bayanus var. *bayanus*, *Saccharomyces* *exiguus*, *Saccharomyces servazzii*, *Zygosaccharomyces rouxii*, *Saccharomyces kluyveri*, *Kluyveromyces thermotolerans*, *Kluyveromyces* *lactis* var. *lactis*, *Kluyveromyces marxianus* var. *marxianus*, *Pichia* *angusta*, *Debaryomyces hansenii* var. *hansenii*, *Pichia sorbitophila*

FEATURES
source
1. 964
Location/Qualifiers
"organism="Candida tropicalis"

```
BASE COUNT      269 a      106 c      78 g      425 t      86 others
ORIGIN
/strain="CBS_94"
/db_xref="taxon:5482"
/clone="XBD0AA002B08"
/clone_1ib="XBD0AA"
/note="end : T7"
```

[illegible]

Db 333 CTRAAAMWTTWGTAAMTGTATGTAAAAAARAAAMRWTTGGMAAAMWMTTAAWAMW 274

QY 1520 ttaacattaaatgcttcctbaatgaatcttcacgaatactccatgctgacattcgaag 1579

Db 273 AAAAAAAAMWTTGTGMAWMAAARAAAAAAMAAWMTTAAAAAAMWTTAARATGCTW 214

QY 1580 agcatatgcttcaaaagaaacaaaaaaa 1610

```
D0      213 WWWTTAAATTTAGAAATAAAAAAAAAAAAAA 183
```

CNS01807/c					
LOCUS					
CNS01807	1101 bp	DNA	linear	GSS 26-JUL-1999	
DEFINITION	Drosophila melanogaster genome survey sequence Sp6 end of BAC				
				accession number U04414	

ACCESSION	fly); genomic survey sequence.
VERSION	AL108721
KEYWORDS	AL108721.1 GI:5629025
SOURCE	GSS.
ORGANISM	fruit fly.
	<i>Drosophila melanogaster</i>
	Neurospora; Trypanosoma; Insecta;

REFERENCE
AUTHORS
TITLE
JOURNAL

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1101)
Genoscope.
Direct Submission
Submitted (23-JUL-1999) Genoscope - Centre National de Séquençage :
BP 191 9106 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr

COMMENT
Determination of this BAC and sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - <http://www.edgp.ebi.ac.uk>. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaut at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector

FEATURES	Location/Qualifiers
source	1..1101
	/organism="Drosophila melanogaster"
	/plasmid="pbelobAc11"